

## De forbedrede nationale test er lanceret,

- **men Beregneren kan ikke længere anvendes efter d. 25. februar 2021**

NordicMetrics har gennem flere projekter i samarbejde med kommuner foretaget statistiske undersøgelser af sammenhængen mellem skalaerne (ækvivaleringer) i de nationale test (DNT). Denne viden blev indbygget i et regnearkbaseret redskab kaldet "Beregneren". I 2013 blev den første udgave lanceret, og den kunne omsætte alle DNT's læsetests resultater til samme intervallskala. På den måde kunne lærerne for første gang opgøre progressionen i dansk, læsning over tid på en skala. Siden blev statistiske undersøgelser af matematiktestene gennemført, hvilket gjorde det muligt at fremstille en udgave af beregneren, så skolerne også kunne opgøre progressionen i matematik. Skalaerne i DNT har været de samme siden lanceringen i 2010, men d. 25. februar 2021 sker der en grundlæggende skalaomlægning, der bevirker, at beregneren ikke fungerer. Samtidigt kommer der nu et ekstraordinært behov for viden om elevernes progression.

I den kommende tid vil der skulle ske en voldsom indsats på de danske skoler for at rette op på det som COVID19 har forvoldt - [600 mio. kr. er afsat til formålet i 2021](#). Der vil afstedkomme et stort behov for at kunne identificere forskellene mellem læring ved fjernundervisning under epidemien og den traditionelle undervisning, både hvad angår trivsel og faglighed. COVID19's effekt på fagligheden kan vurderes ved at se på elevernes progression. Det faglige niveau kan måles validt og reliabelt med DNT. Men for at bruge testene til måling af progression (som med beregneren), skal der foretages helt nye statistiske opgørelser (ækvivaleringer) af de nye skalaers indbyrdes sammenhæng.

### Grundlæggende ændringer i de nationale test 25. februar 2021

De nationale test har især på to områder været kritiseret af testfaglige eksperter i den gennemførte evaluering af DNT: Dels har undersøgelser vist, at opgavernes oprindelige sværhed (målt ved lineære afprøvninger i 2010 og 2014) ikke fuldstændigt afspejler den faktiske aktuelle sværhed af opgaverne i adaptiv drift, og dels er det blevet fremhævet, at testresultaternes statistiske usikkerhed er for stor i forhold til internationale standarder.

Undervisningsministeriets Styrelse for IT og læring (STIL) har som konsekvens af denne kritik genberegnet alle resultater efter de retningslinjer, som blev anbefalet af eksperterne i evalueringen. Om dette skriver Ministeriet på [testogprøver.dk](#): "Den 25. februar forventer Styrelsen for It og Læring at lancere en mindre opdatering af testsystemet bag de nationale test. Opdateringen indebærer en sammenlægning af tilbagemeldingen i testen, således at eleverne får ét samlet resultat for testen i stedet for et resultat pr. profilområde. Opdateringen forbedrer præcisionen af elevernes resultater. Derudover vil elevernes individuelle resultater fra før de obligatoriske test i 2015 fjernes."

Denne "mindre opdatering" betyder massiv forbedring af testenes måleegenskaber, statistiske usikkerhed og reliabilitet: Alle testopgavernes sværheder opdateres, tre skalaer pr fag sammenlægges til én og alle elev-resultater 6 år tilbage i tiden er genberegnet.

**Men på et punkt er der ikke sket noget: Progression måles fortsat ikke i testsystemet.**

Resultaterne i de fag med flere test (engelsk, matematik og dansk, læsning) vil fortsat blive opgjort på forskellige skalaer, der er normeret ud fra gennemsnittet på klassetrinnet, hvor testen er obligatorisk. Så gennemsnittet for hver af de fire læsetest vil være ca. 25 – men 25 i læsning

betyder noget helt forskelligt i 2. og i 8. klasse. Det problem, som Beregneren blev skabt til at løse i 2013 (*at bygge bro mellem testene i faget så man kan måle progression*) er således fortsat ikke adresseret i testsystemet.

<i>Fra 2010 til ult. februar 2021</i>	<i>Primo marts 2021 og frem</i>
Opgavernes sværhed blev beregnet på grundlag af lineære afprøvninger gennemført i 2010/2014, hvor A) der ikke var nogen sammenhæng mellem elevernes dygtighed og opgavernes sværhed, B) eleverne vidste, at ingen ville få resultaterne af testen (intet incitament til at gøre sit bedste).	Opgavernes sværhed estimeres på grundlag af adaptive driftsdata fra 2019, og elevernes dygtighed genberegnes på grundlag af de re-estimerede opgavesværheder. STIL har genberegnet elevresultater så langt tilbage som muligt, dvs. til og med 2015.
Testen er blevet gennemført som tre særskilte test på tre forskellige skalaer (profilområder) med ca. 1/3 af opgaverne på hver skala. Der var typisk 10-25 opgaver pr. pr. profilområde, hvilket medfører høj statistisk usikkerhed (dvs. lav reliabilitet). Samlet resultat blev beregnet som et rå gennemsnit af profilområdescorer.	Testene behandles nu som én test på en skala i stedet for tre, og der opgøres kun ét resultat på grundlag af alle opgaver (typisk 35-75 opgaver pr test). Herved bliver den statistiske usikkerhed meget lav, hvilket betyder meget høj reliabilitet.
Opgavesværheder og elevernes resultater blev rapporteret på Percentil- og Normbaserede skalaer, og elevernes resultater blev endvidere vist på en såkaldt Kriterieskala. Disse er alle ordinalskalaer. Afstandene på skalaerne repræsenterer altså ikke den samme læring forskellige steder på skalaen (ækvivalente).	Resultaterne bliver rapporteret på en intervalskala fra 0-50 efter grundlæggende samme principper som i de internationale undersøgelser (TIMS, PISA, PIRLS osv.) og i de nationale prøver i Norge. Det betyder, at skalaerne bliver ækvivalente – og man kan regne på forskelle i skalaværdier.

Se: <https://www.uvm.dk/folkeskolen/elevplaner-nationale-test--trivselsmaaling-og-sprogproever/nationale-test/om-de-nationale-test>

## Bliver de adaptive nationale test afskaffet snart?

Det blev i efteråret 2019 blev udmeldt de DNT skulle afskaffes og hurtigt erstattes af et ”varigt alternativ”, der, ligesom DNT, kan omfatte alle folkeskoler. Dette alternativ har mange spekuleret over hvad kunne være, men der er endnu ikke præsenteret nogle levedygtige forslag. I mellemtiden har ministeriet og en gruppe forskere evalueret DNT, og ministeriet har fulgt forskernes råd om at forbedre DNT.

De forbedrede adaptive nationale test, der altså lanceres her i marts 2021, er nu ændret på nogle centrale områder, og forskernes mest markante kritikpunkter er udbedret (jf oversigten ovenfor). Samtidigt kan DNT nu noget, som ingen andre danske test formår (de måler med meget lav usikkerhed på elevniveau for alle elever, leverer resultater på intervalniveau, og de er nemme at gennemføre for skolen), og andre lande (fx Norge og Wales) forsøger at etablere lignende systemer. Som Ministeriet tydeligt understreger, kan DNT ikke stå alene, for de kan ikke alt – de skal ses som et supplement til lærerens kvalificerede vurdering og øvrige evaluering. Men det, de måler, gør de bedre end noget andet redskab på dansk grund.

Hvis det i dag blev besluttet, hvad der skulle træde i stedet for DNT, ville det erfaringsmæssigt tage mindst 3 år at udvikle og afprøve det nye system – dette er erfaringsmæssigt meget hurtigt for udvikling af komplicerede nationale måle- og evalueringsredskaber (de nationale test blev eksempelvis taget i brug 6 år efter processen blev sat i gang i 2004). Med udvikling og implementering går der sandsynligvis **mindst 5 år**, før en værdig arvtager til DNT er funktionel, implementeret og klar til drift.

## Beregneren – Baggrund og udbredelse

DNT blev introduceret i 2010 med 10 obligatoriske test i hhv. dansk-læsning, matematik, engelsk og de tre naturfag. Resultaterne fra alle testene blev opgjort på percentil-skalaer fra 1-100 point. Det mest testede område var danske-læsning, hvor der var fire obligatoriske læsetest i hhv. 2., 4., 6. og 8. klasse.

**Baggrund for beregneren:** Hvis en elev scorede 67 point i 4. klasse-testen (altså lå blandt den bedste 33% af landets elever) og to år senere scorede 39 point i 6. klasse-testen (altså nu var blandt den dårligst præsterende 39% af eleverne) kunne man sige, at testresultaterne tydede på, at eleven ikke havde udviklet sine læsefærdigheder lige så meget som normalt fra 4. til 6. klasse.

Man kunne med rimelig sikkerhed sige, at eleven var sakket bagud ift. sine jævnaldrende men man kunne **intet sikkert sige**, om hvorvidt eleven var blevet bedre eller dårligere til at læse i løbet af de to år. For at kunne vurdere læseudviklingen, skulle der "bygges bro" mellem skalaerne, som testresultaterne blev målt på.

Dette løste NordicMetrics gennem et udviklingsarbejde (udført i samarbejde med Københavns Kommune), hvor godt 1.000 elever bidrog ved at tage flere læsetest (fx tog godt 400 elever fra 3. kl. nogenlunde samtidigt en 2. kl. test og en 4. kl test), således at man statistisk kunne opgøre relationerne mellem skalaerne (den statistiske teknik kaldes ækvivalering). Resultaterne blev indbygget i en regnearksløsning som blev kaldt "Beregneren", og som kunne omregne alle resultater i dansk, læsning til én og samme skala. Derved kunne lærerne vurdere eleveres udvikling mellem test uanset hvilke test, der var taget, og om testene var obligatoriske eller frivillige. Beregneren leverede resultaterne på den nye skala fra 0-2000 i tal, men viste også elevernes progression grafisk på en intuitiv og forståelig måde, jf. eksemplerne i denne [vejledning](#). Beregneren til måling af progression i DNT-læsning blev introduceret i 2013 (se [kronik fra 2014](#)), og den blev siden videreudviklet til også at dække matematik.

**Udbredelsen af Beregneren:** Halvdelen (i alt 49) af landets kommuner har fået stillet beregneren til rådighed for deres skoler fra 2013-2020. Af dem har 17 kommuner i perioder indgået i tættere samarbejde med NordicMetrics om at uddanne skoleledere og lærere samt kommunale forvaltningsmedarbejdere. Uddannelsesindsatsen har hovedsageligt bestået af kortere intensive workshops (af ½ til 1 dags varighed) i at forstå, behandle, fortolke og anvende resultater fra DNT i den pædagogiske praksis samt at gennemføre testene, således at testene leverede valide resultater (i den proces spiller læreren en central rolle). I forbindelse med disse workshops blev Beregneren stillet til rådighed som kursusmateriale og omdrejningspunkt som analyseredskab både på og efter workshops.

Dette skete i praksis ved at der (i forbindelse med aftale om at afholde af Workshops i fortolkning og anvendelse af testresultater) blev etableret en hjemmeside specifikt til hver af de 17 kommuners skoler, hvorfra skolens lærere kunne downloade de forskellige udgaver af Beregneren. Disse hjemmesider er løbende blevet opdateret og vedligeholdt, og der har løbende (trods den megen dårlige – og mestendels uretfærdige – omtale) været en lang række skoler/vejledere/lærere, der systematisk har benyttet Beregneren i deres bestræbelser for at for belyse elevernes faglige niveau og udvikling på et videnskabeligt funderet grundlag, og derigennem afsøge handle muligheder for at give en bedre undervisning til eleven og klassen.

## Udvikling af et nyt statistisk grundlag for at bygge bro mellem testene i faget

Regnearksapplikationen ”Beregneren” fungerer it-teknisk og indholdsmæssigt frem til udgangen af februar hvor skalaerne ændres grundlæggende. Når de nye DNT-skalaer introduceres primo marts 2021, vil det statistiske grundlag for beregneren skulle genberegnes, da alle skalaerne er ændret fundamentalt.

**Udvikling af statistisk grundlag:** En ny beregning af det statistiske grundlag for en ny udgave af beregneren (arbejdstitel ”Brobyggeren”) vil kunne ske som en administrativ opgave i samarbejde med et mindre antal kommuner og i alt 10-15 skoler – afhængigt af elevtal.

Det vil kræve udvikling af to brobygningstest (en brobygningstest til 3-6 kl. i matematik og en brobygningstest til 4-7 kl. i engelsk er allerede under udvikling) og koordinering/strukturering af dataindsamlingen samt instruktion/vejledning af de deltagende skoler.

Da det forudgående arbejde med beregneren allerede er sket, ligger der en valideret it-løsning (den seneste udgave af beregneren), der relativt let kan rettes til, så den anvendes på et nyt datagrundlag. Projektet kan derfor gennemføres med en overkommelig ressourceindsats fra kommunens side og med en begrænset ulejlighed for de deltagende skoler.

**Aktivitetsplan:** Metoden til ækvivalering består i, at et antal elever (mindst 250 men gerne 500 elever eller flere) nogenlunde samtidigt og med samme instruktion besvarer to test hver - de test der skal ækvivaleres. Til brug herfor foreslås inddraget et antal skoler med mindst to (og meget gerne flere) spor. Forudsætningen for at kunne beregne sammenhængen mellem de nye skalaer er, at de deltagende skoler lader de relevante elever (eller flest belejligt heraf) gennemføre to test hver i løbet af foråret efter den plan, der fremgår af Tabel 1.

En del af disse test skal skolerne i forvejen gennemføre (de obligatoriske test i matematik, 3. og 6. klasse samt testene i engelsk i 4. og 7. klasse).

Tabel 1: Oversigt over ækvivaleringsarbejdet med de nationale test

Testens emne	Elevers klassetrin	Ækvivalerings test 1	Ækvivalerings test 2
Dansk-Læsning	3 kl.	DNT (frivillig) 2. kl. læsning	DNT (frivillig) 4. kl. læsning
	5 kl.	DNT (frivillig) 4. kl. læsning	DNT (frivillig) 6. kl. læsning
	7 kl.	DNT (frivillig) 6. kl. læsning	DNT (frivillig) 8. kl. læsning
Matematik*	3 kl.	DNT (obligatorisk) 3. kl. matematik	NM Brobygningstest 3-6 kl. mat
	6 kl.	DNT (obligatorisk) 6. kl. matematik	NM Brobygningstest 3-6 kl. mat
	7 kl.	DNT (frivillig) 6. kl. matematik	DNT (frivillig) 8. kl. matematik
Engelsk	4 kl.	DNT (obligatorisk) 4. kl. engelsk	NM Brobygningstest 4-7 kl. engelsk
	7 kl.	DNT (obligatorisk) 7. kl. engelsk	NM Brobygningstest 4-7 kl. engelsk

\* Kan også være 4. og 5. klasse elevers frivillige matematik test til hhv.3. og 6. klasse i kombination med brobygningstest

**Tidsplan:** Skolernes bidrag bliver altså at gennemføre to test pr. fag for en række af deres elever. Tidsplanen kan tilpasses omstændighederne (udviklingen i Corona, smitteudvikling og restriktioner i Danmark), men det skønnes, at det uden problemer vil kunne nås at indsamle de nødvendige data i dette skoleår, også selvom skolerne først åbner for alle klassetrin sent i foråret. Dybest set er kravet til en klasses deltagelse en indsats af en varighed på 2 lektioner – med en vis forberedelse for læreren (dog næppe så meget som almindelige undervisning kræver).

## Ekstra vigtigt at skolerne får adgang til aktuelle data efter COVID19

Nedlukningen under COVID19 har betydet, at en stor del af eleverne har været isoleret hjemme sammen med familierne. Der er mange elever og familier, der har kunnet håndtere nedlukningen uden sociale- og psykiske belastninger for eleverne og uden betydelige læringstab. Men der er mange tegn og undersøgelser, der kunne pege i retning af at belastningen ved COVID19 har en stærk social og etnisk slagside.

Ny europæisk forskning antyder, at det faglige COVID19-tab er voldsomt<sup>1</sup>. Middelklassefamilier har været i stand til at opretholde en god uddannelses- og social kvalitet for deres børn, mens børn fra socialt udsatte familier har haft dårlige muligheder både med hensyn til tid og læringserfaringer (skolearbejde og vedligeholdelse af fritidsaktiviteter).

**Lighedspunkter med sommerlæringstab:** Den amerikanske skoleforskning har i årtier arbejdet med begrebet "Summer Loss": At eleverne – især de elever fra socialt udsatte hjem - har tendens til at opleve "sommerindlæringstab", et fald i præstationer mellem forår og efterår, der udvider kløften mellem socialt stærke og svage elever.

Erfaringer fra USA er, at middel- og højindkomst kan kompensere for den negativ effekt af ikke at gå i skole, mens udsatte familiers børn klarer sig dårligere. Og læringstabet indhentes ikke efter ferien så forskellene ("the gap") udvides over tid.

Nyere amerikansk forskning<sup>2</sup> viser yderligere, at læringstabet er komplekst. Analyser af data fra de 3,4 millioner elever, der går på skoler, som anvender [NWEA's](#) MAP Growth-vurderinger, viser,

at sommertabet kan undgås. NWEA er den største amerikanske forskningsbaserede non-profit udbyder af adaptive test (meget lig den nye udgave af DNT) til brug for pædagogisk udvikling, der måler fagligt niveau og progression efter samme principper som DNT/Beregneren. Men denne nyere forskning sætter også spørgsmålstegn ved de sædvanlige midler for at styrke det faglige niveau, såsom adgang til ekstraundervisning og forlængelse af skoleåret. Kun få ting er forskere og praktikere enige om, og to af disse er: 1) Effekten af fx COVID på fagligt niveau og udvikling har en markant social slagside, 2) Skolerne kan gøre noget ved det, hvis de har den nødvendige viden om identifikation af problemer og handlemuligheder.

Forskning (primært fra de engelsktalende lande) viser altså, at viden om problemet (hvilke elever har udfordringer på hvilke områder) og viden om de rette pædagogiske redskaber begge dele er nødvendige for at kunne handle effektivt.

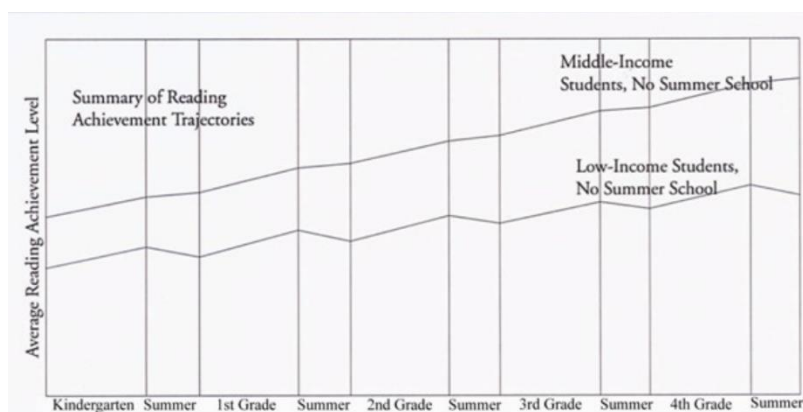


Illustration af sommerlæringstab for middel- og lavindkomstfamilier (Kilde: Målinger foretaget af Metametrics)

<sup>1</sup> Bonal, X., & González, S. (2020). The impact of lockdown on the learning gap: family and school divisions in times of crisis. *International Review of Education*, 1-21.

<sup>2</sup> Kuhfeld, M. (2019). Surprising new evidence on summer learning loss. *Phi Delta Kappan*, 101(1), 25-29.

**DNT kan bruges til at identificere COVID19-effekten:** Faglig problemer kan identificeres som lav eller endog negativ progression i løbet af COVID19-epidemien. En opgørelse forudsætter altså en før- og eftermåling på én skala, der er egnet til måling af progression. Hvis DNT suppleres med Brobyggeren kan det blive et sådant redskab.

Der er stor faglig spredning i elevernes faglighed. I nogle lande (også europæiske fx Holland) kan de dominerende forskelle henføres til skolernes faglige gennemsnit. Danmark er et af de lande i verden, hvor der er mindst forskel indbyrdes mellem skolernes faglige gennemsnit. Kun 10-15% af den faglige spredning, der kan henføres til skolerne og klassen. De resterende 85-90% af spredning konstateres mellem eleverne internt i klassen<sup>3</sup>.

Det betyder, at der sandsynligvis vil være én eller et par elever i langt de fleste klasser, der kan have særlige negative faglige følgevirkninger af COVID19, som kan identificeres som lav eller negativ progression, og som der kan gøres noget ved, hvis det opdages. Og det behøver ikke altid være de lavest præsterende elever.

Skolerne bør have en særlig opmærksomhed på de elever, der kommer fra hjem, hvor det sprog, der tales til dagligt, ikke er dansk, og de elever som kommer fra hjem, hvor der ikke er den optimale opbakning og hjælp til skolearbejdet. I perioder med hjemmeundervisning har der været ekstra behov for, at der øves, lyttes og tales dansk i disse hjem. Dette er mange skoler meget opmærksomme på<sup>4</sup>, men det er langt fra alle elever, der har fået tilstrækkelig støtte og træning – det beror blandt andet på en lang række faktorer (fritidsaktiviteter, kammerater, forældres indsats mv.). Og disse elever har skolen ekstra vanskeligt ved at følge under en periode med nedlukning.

Næsten alle lærere i folkeskolen, der underviser i testfagene, kan få behov for hjælp til at identificere elever, der har ekstra brug for et løft efter COVID19.

De bedste data til at understøtte lærernes vurderinger vil være valide og reliable mål for elevernes progression over en længere periode. Og når det drejer sig om matematik og læsning vil DNT kunne levere en unik viden til skolen og den enkelte lærer. For at kunne bruge resultaterne til at vurdere progression vil der være et stort behov for den efterbehandling, der hidtil har været leveret af beregneren. Det behov vil "brobyggeren" kunne dække fremadrettet.

**Yderligere udvikling:** Der arbejdes på at etablere samarbejder med forskning og udvikling, herunder et samarbejdsprojekt med Professionshøjskolen Absalon, som bl.a. sigter på at udnytte de data (i anonymiseret form), der kommer ud af afprøvningsresultaterne i en forskningsmæssig kontekst. Projektet har yderligere til formål at udvikle og afprøve innovative ideer til videreudvikling af anvendelsen af DNT (med eller uden Brobyggeren) gennem inddragelse af skoler, lærere og lærerstuderende.

JW 21/04 2021

<sup>3</sup> Christensen, V. T. (red.), (2019). PISA 2018 - Danske unge i en international sammenligning, VIVE

<sup>4</sup> Se fx <https://skoleliv.dk/debat/art8073240/M%C3%A5neders-sprogarbejde-tabes-n%C3%A5r-skolen-er-lukket>